

Pulsed Bias Etching of Silicon Pillar Arrays: Experiment and Surface-Memory Simulation of Loading Control

Jianming Guo¹, Zhengming Liu², Mingqiang Geng¹, Kunrun Song², Huiyun Jiang^{1,2,*}, Dong Ni^{1,*}, and Dawei Gao^{1,2}

¹College of Integrated Circuits, Zhejiang University, Hangzhou 311200, China

²Zhejiang ICsprout Semiconductor Co., Ltd., Hangzhou 311200, China

Email: Jianming Guo(21826022@zju.edu.cn)

* Corresponding author: dni@zju.edu.cn, huiyun.jiang@zju.edu.cn

Abstract—Silicon pillar arrays are increasingly used in emerging nanophotonic and metasurface devices, where wafer-scale fabrication demands tight control of etch-depth uniformity and profile fidelity. Their etching is strongly affected by aspect-ratio-dependent etching (ARDE), loading-induced depth nonuniformity, and bottom-profile instability. We combine pulsed-bias etching experiments with a GPU-accelerated transient feature-scale simulation to study duty-cycle-dependent loading control. A surface-memory model couples Monte Carlo particle transport, ion–neutral synergistic etching, and transient surface coverage and byproduct blocking, with a secondary post-interface charge correction. Enabled by GPU-parallel particle tracing ($>2.6 \times 10^6$ trajectories/s), a full transient etch completes in under one minute on a single GPU, making systematic duty-cycle scans practical. Simulations agree with experiments and show that pulsed bias reduces ARDE by periodically depleting and recovering the local reactive surface state, providing guidance for duty-cycle optimization.

Keywords—plasma etching, ARDE, feature-scale simulation, pulsed bias, metasurface

I. INTRODUCTION

Wafer-scale metasurface fabrication requires high-fidelity etching of silicon pillar arrays with coexisting pattern densities and critical dimensions (CDs). The pillar height, CD, sidewall angle, and bottom profile directly determine the optical response of the final device. However, when isolated and dense patterns are etched simultaneously, local differences in open area and feature geometry lead to nonuniform particle transport, ion access, neutral supply, and mask consumption. As a result, different regions can advance at different rates, producing aspect-ratio-dependent etching (ARDE), loading-induced depth nonuniformity, and profile instability[1,2].

Under continuous-wave (CW) bias, this imbalance becomes particularly severe because the surface is continuously exposed to ion bombardment. Isolated or small-CD regions can etch faster, lose hard-mask integrity earlier, and collapse before larger-CD or denser neighboring patterns reach the target depth. Pulsed bias provides a possible route to mitigate this failure by separating ion-driven removal and surface recovery in time. However, reducing the duty cycle also lowers the effective vertical etch rate, so profile improvement must be balanced against throughput loss and possible lateral degradation[3,4].

In this work, we combine pulsed-bias silicon pillar etching experiments with a GPU-accelerated transient feature-scale simulation framework to study duty-cycle-dependent loading control. The simulator couples Monte Carlo particle transport, ion–neutral synergistic etching, transient surface coverage, byproduct blocking, and a secondary post-interface charge correction. Unlike steady-state or purely time-averaged descriptions, the model retains the local surface state as a memory variable, so that the etch rate depends on the previous ON/OFF history of the surface.

II. EXPERIMENTAL

Fig. 1 shows the wafer-scale metasurface silicon pillar pattern and the corresponding etching structure. The wafer contains coexisting dense and isolated pattern regions with different critical dimensions (CDs), which provides a direct platform for evaluating pattern-dependent loading and ARDE behavior. The structure consists of silicon pillars on a SiO₂ underlayer, patterned through a photoresist/BARC/APF hard-mask stack. Etching was performed in an ICP system using SF₆/CF₄/CHF₃/O₂-based chemistry under continuous-wave (CW) bias and pulsed-bias conditions with duty cycles of 50%, 30%, and 10%.

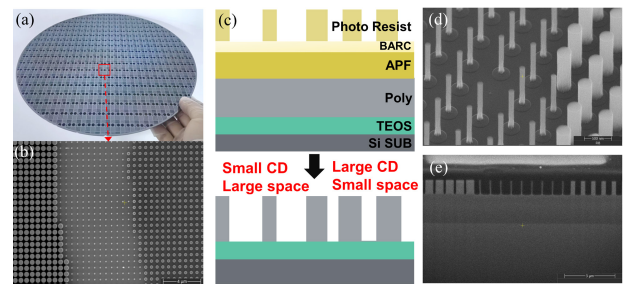


Fig. 1. Wafer-scale metasurface silicon pillar etching overview. (a) Wafer-level pattern distribution. (b) Local region with different densities and CDs. (c) Schematic of the mask/film stack and target silicon pillar geometry. (d,e) Representative SEM images of etched silicon pillar arrays.

The first experimental observation is the existence of strong ARDE/loading behavior across regions with different CDs and pattern densities. The etch advancement is not uniform across the wafer pattern, indicating that local feature geometry and open area strongly affect the etching response. Under CW bias, this nonuniformity develops into a clear profile-failure mode. As shown in Fig. 2, isolated small- and medium-CD pillars

collapse earlier, whereas neighboring denser patterns remain comparatively intact.

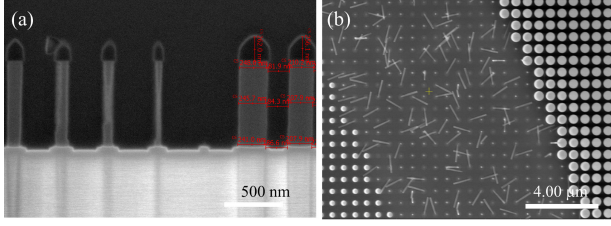


Fig. 2. CW etch results. (a) Cross-sectional view. (b) Tilted views showing earlier collapse of isolated small- and medium-CD pillars.

The second observation is the duty-cycle dependence of the effective etch rate and profile morphology. Compared with CW bias, pulsed bias suppresses the collapse tendency of isolated pillars. However, the vertical etch rate decreases as the duty cycle is reduced from 50% to 30% and 10%. Therefore, lower duty cycle improves profile stability in some regions, but also reduces vertical propagation and may require a longer etch time to reach the target depth. Fig. 3 summarizes the pulsed-bias etching results. At 50% duty cycle, the pillars remain well defined and show a relatively balanced profile. At 30% duty cycle, stronger taper is observed, indicating more pronounced sidewall angle degradation. At 10% duty cycle, the vertical etch rate is further reduced; collapse is suppressed, but lateral consumption becomes more evident because of the longer etching time.

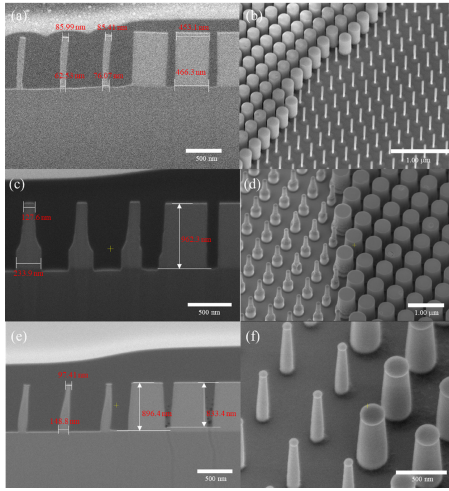


Fig. 3. Pulsed bias results at different duty cycles. (a,b) 50% duty. (c,d) 30% duty. (e,f) 10% duty.

In addition to depth nonuniformity and taper, bottom-profile instability is also observed near the SiO₂ interface. Regions that reach the interface earlier are more susceptible to bottom narrowing or notch-like distortion during over-etch. This phenomenon is treated in the simulation as a post-interface effect and is discussed later together with the secondary oxide-charge correction. Overall, the experiments reveal three coupled trends: ARDE/loading-induced nonuniform etch advancement, duty-cycle-dependent etch-rate reduction, and bottom-profile instability after local SiO₂ exposure.

III. MODELING FLOW

To interpret the duty-cycle-dependent etching behavior of silicon pillar arrays, we developed a GPU-accelerated transient feature-scale simulator. The simulator is a reduced-order feature-scale model rather than a chamber-scale plasma solver. Its purpose is to capture the coupled evolution of local particle transport, ion-neutral synergistic surface reactions, transient surface states, and etched profiles under continuous-wave (CW) and pulsed-bias operation.

The process and profile inputs include the duty cycle, ON/OFF bias waveform, ion energy and angular distribution, ion and neutral fluxes, and the initial mask/Si/SiO₂ stack geometry. The feature structure is mapped onto a two-dimensional voxel domain containing multiple openings with different widths. Each voxel stores a material identifier and a continuous occupancy value. During etching, the occupancy is gradually reduced according to the local removal rate and is converted to vacuum once it falls below a prescribed threshold. This voxel-based representation allows the simulator to update the evolving profile and the local surface state within the same computational framework.

The simulation proceeds through an explicit transient loop. At each time step, an RF-state scheduler determines whether the system is in the ON or OFF state. CW operation is treated as continuously ON. For pulsed bias, each ON/OFF step represents an RF-cycle-averaged condition rather than an individual RF oscillation. The RF state controls the ion probability, ion energy, ion angular spread, neutral sticking behavior, surface recovery, and byproduct purge. A large batch of ions, neutrals, and byproduct particles is then traced by GPU-accelerated Monte Carlo ray marching. Particle incidence, reflection, adsorption, sputtering, and redeposition events are accumulated and used to update the surface reaction and profile evolution. GPU acceleration is necessary because statistically meaningful local fluxes require many particle trajectories over many transient steps. On a single GPU, this implementation traces over 2.6×10^6 particle trajectories per second, so that a complete transient etch ($\sim 10^8$ trajectories over several thousand time steps) finishes in under one minute, making systematic duty-cycle scans computationally practical.

The core of the simulator is the surface-memory reaction kernel. Reactive neutrals adsorb on the silicon surface and build an active surface coverage, denoted as (θ) , while ion bombardment activates a physical sputtering channel and a dominant ion-neutral synergistic removal channel. The local neutral supply is not assumed to be spatially uniform. Instead, the effective neutral flux arriving at each local surface, (J_n^{eff}) , is updated according to the particle transport statistics in the evolving feature geometry. For interpretation, an effective transmission factor can be defined as $(T_{\text{cl}} = J_n^{\text{eff}} / J_{n,0})$, where $(J_{n,0})$ is the nominal incident neutral flux at the feature entrance. Therefore, aspect-ratio-dependent loading is not imposed directly on the etch rate, but enters through the local neutral supply that controls the active coverage (θ) .

The local silicon etch rate is then evaluated from the instantaneous surface state. Following the Coburn-Winters ion-neutral synergy concept, the removal rate consists of a small

physical sputtering contribution and a dominant synergistic contribution controlled by the active coverage (θ):

$$R = \left[Y_{\text{phys}} + Y_{\text{syn}} \frac{\theta}{\theta + K} \right] E f(\alpha) \exp(-\lambda\beta) \quad (1)$$

Here, R is the local silicon etch rate, (Y_{phys}) and (Y_{syn}) are the physical and synergistic yield coefficients, respectively, (E) is the RF-cycle-averaged ion energy, ($f(\alpha)$) is the angular yield function, (K) is the half-saturation constant of the synergistic channel, and (β) is the byproduct blocking factor. The exponential term describes the suppression of etching caused by local blocking or redeposition. In this formulation, the feature-scale transport effect does not directly multiply the etch rate; instead, it modifies (J_n^{eff}), which controls (θ) and therefore changes the synergistic removal channel.

Unlike steady-state feature-scale models, the active coverage (θ) and the blocking factor (β) are retained as transient surface-memory variables. Their evolution is written as:

$$\frac{d\theta}{dt} = k_{\text{ads}} J_n^{\text{eff}} (1 - \theta) - k_{\text{des}} \theta - k_{\text{cons}} J_i \theta \quad (2a)$$

$$\frac{d\beta}{dt} = k_{\text{gen}} R - k_{\text{purge}} \beta \quad (2b)$$

Here, J_i is the local ion flux, while k_{ads} , k_{des} , k_{cons} , k_{gen} , and k_{purge} are effective surface-state kinetic coefficients. During the ON phase, directional ion bombardment consumes the active coverage and promotes ion-assisted material removal and byproduct generation. During the OFF phase, the reduced ion flux allows neutral adsorption to restore the active coverage, while the blocking factor is partially purged. Therefore, the local reaction rate depends not only on the instantaneous particle fluxes, but also on the previous ON/OFF history of the surface. This delayed, history-dependent response is referred to as surface memory.

Pulsed bias is coupled to the simulator through the RF-state scheduler rather than by simply scaling the average ion energy. For each time step, the scheduler determines whether the local plasma condition corresponds to the ON or OFF portion of the bias waveform. The ON/OFF state changes the ion probability, ion energy, ion angular spread, neutral sticking behavior, surface recovery, and byproduct purge. In this way, two recipes with similar average bias dose can still produce different profiles because they generate different histories of (θ), (β), and local particle incidence.

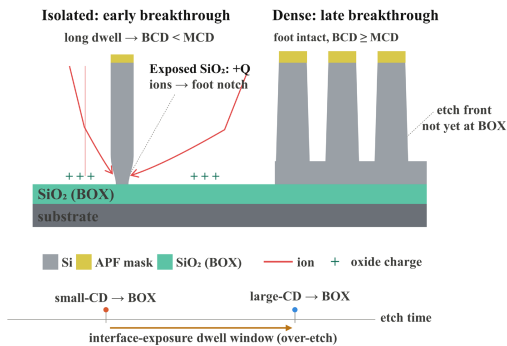


Fig. 4. Schematic of charge-assisted bottom notching after local BOX exposure.

A secondary oxide-charge variable (Q) is introduced to describe the post-interface bottom-profile evolution. This variable is activated only after the local etch front reaches the SiO_2 buried oxide (BOX). Before BOX exposure, the silicon profile evolution is governed mainly by particle transport and the transient surface-memory variables (θ) and (β). Once the BOX is locally exposed, ion impacts on the oxide increase (Q), while OFF-phase electron neutralization relaxes it. The resulting charge state modifies near-bottom ion trajectories and enhances ion steering toward the pillar foot region.

This charge-assisted correction is used to capture the onset of bottom-corner attack and notch formation during over-etch. As illustrated in Fig. 4, small-CD isolated features reach the BOX earlier than large-CD dense features. The time interval between these two breakthrough events forms an interface-exposure dwell window. During this window, the oxide beneath the isolated feature remains exposed for a longer time, allowing charge accumulation and repeated ion deflection near the foot region. This promotes local lateral attack and leads to bottom notching, which is reflected by a reduced bottom critical dimension (BCD) compared with the middle critical dimension (MCD), i.e., $\text{BCD} < \text{MCD}$. In contrast, dense features reach the BOX later and experience a shorter oxide-exposure time; therefore, the charge-assisted bottom attack is weaker, and the foot region remains relatively intact, often giving $\text{BCD} \geq \text{MCD}$.

IV. RESULTS AND DISCUSSION

Fig. 5 shows the simulated surface coverage θ at the etch front under CW and pulsed-bias conditions. Under CW operation, θ approaches a quasi-steady state, but the steady level is lower for narrower openings because the local neutral supply is more limited. As a result, the synergistic term $\theta/(\theta+K)$ becomes CD-dependent, which leads to stronger loading and a more pronounced ARDE effect. In contrast, under pulsed bias, θ exhibits a periodic sawtooth-like response: it decreases during the ON phase due to ion-assisted consumption and recovers during the OFF phase through neutral adsorption. This repeated ON-consumption/OFF-recovery cycle periodically restores all openings toward a similar coverage range near the Coburn-Winters saturation regime, thereby reducing the CD-dependent variation in the synergistic etching channel. As the duty cycle decreases, the difference in θ among different openings becomes progressively smaller, indicating that pulsed bias weakens the feature-size dependence of the surface reaction.

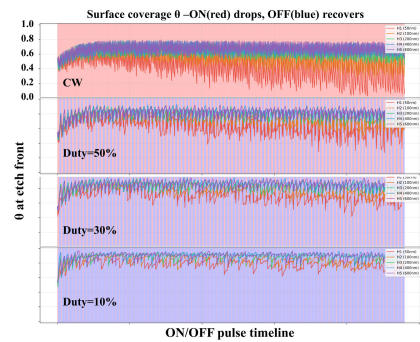


Fig. 5. Simulated surface coverage θ at the etch front under CW and pulsed bias (duty = 50%, 30%, 10%).

The simulated profile evolution is shown in Fig. 6. Under CW bias, the different openings etch at noticeably different rates, and the narrow openings propagate faster than the wider openings, resulting in clear depth nonuniformity. With pulsed bias, the evolution of different openings becomes more synchronized, and the depth difference among CDs is gradually reduced. Quantitatively, the ARDE ratio, defined as the etched depth of the narrowest opening divided by that of the widest opening, increases from 0.706 under CW to 0.805, 0.833, and 0.966 at duty cycles of 50%, 30%, and 10%, respectively. Since the ratio approaches unity, this indicates that the magnitude of ARDE is progressively reduced as the duty cycle decreases.

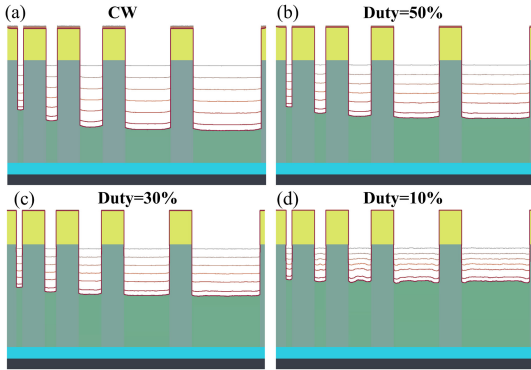


Fig. 6. Simulated etch profiles at CW, 50%, 30%, and 10% duty cycle.

At the same time, the maximum etched depth decreases from 616 nm under CW to 508 nm, 452 nm, and 321 nm for duty cycles of 50%, 30%, and 10%, respectively. This means that although a lower duty cycle improves loading uniformity, it also reduces the effective vertical etch rate. This simulated trend is consistent with the experimental observation that the etch rate decreases monotonically from CW to lower-duty pulsed conditions, while the ARDE effect is simultaneously suppressed. In other words, pulsed bias improves CD-uniformity at the expense of throughput. Among the simulated cases, the 50% duty cycle provides the most balanced behavior, retaining a relatively large etch depth while already showing a clear reduction in ARDE. Further reduction to 30% and 10% continues to improve depth uniformity, but the vertical etch rate decreases more significantly, which would require a longer total etch time in practice and may aggravate lateral exposure or profile degradation.

Fig. 7 compares the final simulated etched profiles. The CW case shows the strongest CD-dependent depth difference and the most evident bottom-profile distortion near the Si/SiO₂ interface, especially in the narrow openings. This behavior is attributed to the combined effect of early interface exposure and the secondary oxide-charge correction. Because small-CD openings reach the BOX earlier than large-CD openings, they experience a longer interface-exposure dwell window during over-etch. During this dwell period, the silicon foot remains exposed to continued ion bombardment, making the bottom

region more susceptible to lateral attack. Once the oxide is locally exposed, ion impacts on the SiO₂ surface further accumulate the charge variable (Q), which modifies near-bottom ion trajectories and enhances ion steering toward the pillar foot.

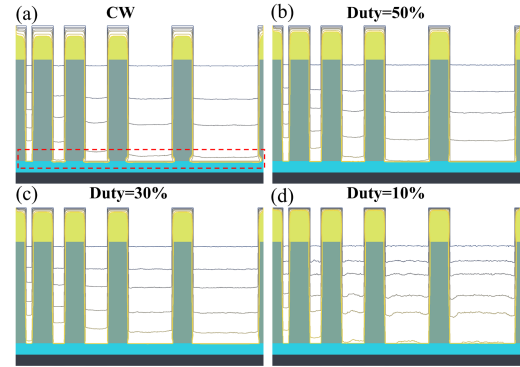


Fig. 7. Final simulated etched profiles under CW and pulsed-bias conditions.

V. CONCLUSION

This work combined pulsed-bias etching experiments with a GPU-accelerated surface-memory feature-scale simulation to investigate loading evolution in wafer-scale silicon pillar arrays. The experiments show clear ARDE/loading behavior, duty-cycle-dependent etch-rate reduction, and bottom-profile instability after local SiO₂ exposure. The simulation reproduces the main experimental trends: decreasing the duty cycle suppresses ARDE and improves CD-dependent depth uniformity through periodic recovery of the active surface state, while also lowering the effective vertical etch rate. The secondary post-interface charge correction further explains the enhanced bottom attack after early oxide exposure. These results demonstrate that the proposed surface-memory simulator captures the experimentally observed duty-cycle trends and provides a physically interpretable tool for analyzing loading evolution in pulsed-bias silicon pillar etching.

REFERENCES

- [1] J. W. Coburn and H. F. Winters, "Ion- and electron-assisted gas-surface chemistry—An important effect in plasma etching," *J. Appl. Phys.*, vol. 50, no. 5, pp. 3189–3196, 1979, doi: 10.1063/1.326355.
- [2] R. J. Belen, S. Gomez, D. Cooperberg, M. Kiehlbauch, and E. S. Aydil, "Feature-scale model of Si etching in SF₆/O₂ plasma and comparison with experiments," *J. Vac. Sci. Technol. A*, vol. 23, no. 5, 2005, doi: 10.1116/1.2013317.
- [3] C. M. Huard, Y. Zhang, S. Sriraman, A. Paterson, and M. J. Kushner, "Role of neutral transport in aspect ratio dependent plasma etching of three-dimensional features," *J. Vac. Sci. Technol. A*, vol. 35, no. 5, 05C301, 2017, doi: 10.1116/1.4973953.
- [4] T. Kinoshita, M. Hane, and J. P. McVittie, "Notching as an example of charging in uniform high density plasmas," *J. Vac. Sci. Technol. B*, vol. 14, no. 1, pp. 560–565, 1996, doi: 10.1116/1.588431.